Taksiran Parameter Multinomial Logit dengan Menggunakan Generalized Method of Moment

Nurfathanah
Mahasiswa Prodi Studi Statistika Universitas Hasanuddin

Erna Tri Herdiani
Mahasiswa Prodi Studi Statistika Universitas Hasanuddin

Georgina Maria Tinungki
Mahasiswa Prodi Studi Statistika Universitas Hasanuddin

ABSTRAK, Regresi logistik multinomial merupakan perluasan dari regresi logistik biner yang memungkinkan lebih dari dua kategori variabel dependen. Pada paper ini akan membahas penaksiran parameter regresi logistik multinomial melalui *Generalized Method of Moment (*GMM). *Generalized Method of Moment* (GMM) merupakan salah satu metode yang dapat mengatasi pelanggaran asumsi pada data seperti autokorelasi dan heteroskedastisitas. Hasil taksiran parameter regresi logistik multinomial logit dengan GMM diperoleh $\hat{\beta} = (X^t ZWZ^t X)^{-1} X^t ZWZ^t Y$.

Kata Kunci*: Regresi Multinomial Logit, Generalized Method of Moment, Momen, Regresi*

1. PENDAHULUAN

Model regresi logistik merupakan salah satu model untuk menganalisis data kategori. Regresi logistik terbagi menjadi dua berdasarkan responnya yaitu regresi logistik biner dan multinomial. Regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data dikotomi/biner dengan variabel bebas yang berupa data berskala interval atau kategori. Model regresi logistik biner dapat diperluas menjadi model regresi multinomial jika variabel respon memiliki lebih dari dua nilai kategori [2].

Penaksiran parameter regresi logistik multinomial dapat dilakukan dengan beberapa metode salah satunya adalah *Generalized Method of Moment (*GMM). Metode GMM diperkenalkan oleh Hansen pada tahun 1982 sebagai estimasi parameter yang meminimalkan bentuk kuadrat dari kondisi momen sampel yang terboboti. Bagi para ahli teori, keuntungan utama GMM adalah menyediakan kerangka kerja yang sangat umum untuk mempertimbangkan masalah-masalah inferensi statistik [3]. Selain itu, Metode *Generalized Method of Moment* (GMM) merupakan salah satu metode yang dapat mengatasi pelanggaran asumsi pada data seperti autokorelasi dan heteroskedastisitas [5].

Adapun tujuan dari penelitian ini adalah untuk menaksir parameter dari model multinomial logit dengan menggunakan *Generalized Method of Moment* (GMM).

2. TINJAUAN PUSTAKA

Model Multinomial Logit

Model linear diperumum merupakan suatu metode membentuk model yang digunakan untuk menyelidiki hubungan antara variabel prediktor dan variabel respon. Metode ini pertama kali diperkenalkan oleh Nelder dan Weddeburn (1972) untuk menentukan taksiran model linear dengan asumsi model error tidak harus berdistribusi normal, tetapi distribusinya termasuk dalam keluarga eksponensial [4].

Menurut Agresti [1] model linear diperumum memiliki tiga komponen yaitu :

1. Komponen random terdiri dari variabel respon $y_i = 1,2, \ldots, n$ yang memiliki hubungan saling bebas dan n menyatakan ukuran sampel. Komponen ini memiliki mean μ_i dengan peluang densitasnya termasuk keluarga eksponensial.

   Misalkan :

   $$Y = \left(y_{1,} y_2, \ldots., y_N\right)^t$$

   Dimana *N* adalah banyaknya observasi dan Y saling bebas.

   $y_i \sim f(y_{i,} \theta_i)$ dimana *i* = 1, 2,..., *N*

   Jika fungsi peluang y_i dapat ditliskan sebagai

   $$a(\theta_i) b(y_i) \exp\left(y_i Q(\theta_i)\right) \qquad (1)$$

Maka Y disebut keluarga eksponensial (*exponential family*)

2. Komponen sistematik yang terdiri dari variabel prediktor x_i, dengan $x_i(1, x_{i1}, x_{i2}, \dots, x_{ik})^t$ dimana k menyatakan banyaknya prediktor. Komponen ini menjelaskan prediktor $\{x_i\}$ dalam suatu persamaan :

$$\eta_1 = \beta_0 + \beta_{i1} + \cdots + \beta_k x_{ik} \tag{2}$$

Kombinasi linear ini disebut prediktor linear.

3. Link adalah suatu fungsi yang menghubungkan antara komponen sistematik dengan mean dari komponen. Misalkan $E(Y_i) = h(\mu_i) = \eta_i$ dimana h disebut link dengan syarat h merupakan fungsi dari μ yang monoton dan differensiable.

Tabel 1. Tabel Frekuensi *N*

Populasi	1	...	j	...	r	
1	n_{11}	...	n_{1j}	...	n_{1r}	n_1
⋮	⋮		⋮		⋮	⋮
i	n_{i1}	...	n_{ij}	...	n_{ir}	n_i
⋮	⋮		⋮		⋮	⋮
s	n_{s1}	...	n_{sj}	...	n_{sr}	n_s

Dimana n_{ij} menyatakan banyaknya sampel dari populasi ke i yang termasuk kategori ke j dengan i = 1, 2, ..., s dan j = 1, 2,.., r.

Distribusi multinomial termasuk keluarga eksponensial. $n_{ij} \sim$ multinomial (n, π_{ij}) dengan i = 1,2,...,s dan j = 1,2,...,r.

Fungsi peluang n_{ij} dapat dituliskan sebagai :

$$a(\pi_{ij}) = \sum_{j=1}^{r} \pi_{ij} = 1$$
$$b(n_{ij}) = \frac{n_i!}{\prod_{j=1}^{r} n_{ij}!}$$
$$Q(\pi_{ij}) = \ln(\pi_{ij})$$

Selanjutnya, misalkan P adalah tabel frekuensi relatif dengan elemen P_{ij} yang merupakan relatif respon Y kategori j pada populasi i, dimana $P_{ij} = \frac{n_{ij}}{n_i}$

Tabel 2. Tabel Frekuensi Relatif P

Populasi	1	...	j	...	r	
1	P_{11}	...	P_{1j}	...	P_{1r}	P_1
⋮	⋮		⋮		⋮	⋮
i	P_{i1}	...	P_{ij}	...	P_{ir}	P_i
⋮	⋮		⋮		⋮	⋮
s	P_{s1}	...	P_{sj}	...	P_{sr}	P_s

Model dengan asumsi respon multinomial dengan link generalized logit disebut model multinomial respon yang kemudian dikenal dengan Model Multinomial logit. Bentuk umum model multinomial logit adalah :

$$\ln \frac{\pi_{ij}}{\pi_{ir}} = X_{ij}\beta \tag{3}$$

Generalized Method of Moment

Metode *Generalized Method of Moment* (GMM) merupakan salah satu metode yang dapat mengatasi pelanggaran asumsi pada data seperti autokorelasi dan heteroskedastisitas. Metode ini diperkenalkan pertama kali oleh Hansen pada tahun 1982 yang didefenisikan sebagai metode estimasi parameter yang hanya tergantung pada kondisi momen yang digunakan.

Secara umum regresi multinomial logit dituliskan dengan persamaan sebagai berikut :

$$Y = X_i\beta + \varepsilon_i \tag{4}$$

Persamaan (4) diasumsikan mengandung variabel instrumen Z_i dengan Z_i adalah sebagian atau keseluruhan dari variabel eksplanatori (X_i) yang tidak berkorelasi dengan (ε_i) sebagai berikut :

$$Z_iY_i = Z_iX_i\beta + Z_i\varepsilon_i \quad i = 1,2,\dots,N \tag{5}$$

Karena Z_i tidak berkorelasi dengan ε_i maka ditulis $E(Z_i\varepsilon_i) = 0$

3. HASIL DAN PEMBAHASAN

Metode GMM didefenisikan sebagai suatu estimasi parameter yang meminimumkan bentuk kuadrat dari kondisi momen sampel data regresi yang terboboti, sehingga

$$S(\hat{\beta}) = [(Z^t\varepsilon)^t W(Z^t\varepsilon)] \quad (6)$$

$$= \left[\left(Y^t - X^t\hat{\beta}^t\right)Z\right]W\left[\left(Z^t\left(Y - X\hat{\beta}\right)\right)\right]$$

$$= Y^tZWZ^tY - Y^tZWZ^tX\hat{\beta} - \hat{\beta}^tX^tZWZ^tY + \hat{\beta}^tX^tZWZ^tX\hat{\beta} \quad (8)$$

Karena $Y^tZWZ^tX\hat{\beta}$ merupakan matriks berordo 1 x 1 maka transposenya yaitu $(Y^tZWZ^tX\beta)^t = \beta^tX^tZWZ^tY$ juga merupakan matriks berordo 1 x 1 yang sama maka

$$S(\hat{\beta}) = Y^tZWZ^tY - 2(\beta^tX^tZWZ^tY) + \beta^tX^tZWZ^tX\beta \quad (9)$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2(X^tZWZ^tY) + 2\left(X^tZWZ^tX\hat{\beta}\right) \quad (10)$$

$$\hat{\beta} = (X^tZWZ^tX)^{-1}X^tZWZ^tY \quad (11)$$

Untuk mengetahui apakah GMM merupakan estimator yang baik, akan ditunjukkan bahwa hasil estimasi GMM bersifat BLUE (*best, Linear, Unbias Estimator*).

***Unbias* (tidak bias)**

$$E(\hat{\beta}) = E(X^tZWZ^tX)^{-1}X^tZWZ^tY$$

$$= (X^tZWZ^tX)^{-1}X^tZWZ^t(X\beta + \varepsilon)$$

$$= (X^tZWZ^tX)^{-1}X^tZWZ^t(X\beta + \varepsilon)$$

$$= \beta \quad (12)$$

Jadi $E(\hat{\beta}) = \beta$ maka $\hat{\beta}$ adalah estimator yang merupakan penaksir tak bias.

3.2. Linear

$$\hat{\beta} = E[X^tZWZ^tX]^{-1}X^tZWZ^tY$$

$$= (X^tZWZ^tX)^{-1}X^tZWZ^tX\beta + (X^tZWZ^tX)^{-1}X^tZWZ^t\varepsilon$$

$$= \beta + (X^tZWZ^tX)^{-1}X^tZWZ^t\varepsilon \quad (13)$$

Merupakan fungsi linear dari β dan ε

Variansi minimum

$$\text{Var}\left(\hat{\beta}\right) = E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^t\right]$$

$$\hat{\beta} = (X^tZWZ^tX)^{-1}X^tZWZ^tY \ (*)$$

$$\hat{\beta} - \beta = (X^tZWZ^tX)^{-1}X^tZWZ^t\varepsilon \ (**)$$

Sehingga jika disubstitusikan ke dalam persamaan matriks variansi dari $\hat{\beta}$ diperoleh:

$$\text{Var}\left(\hat{\beta}\right) = E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^t\right] \quad (***)$$

Subtitusi Persamaan (*) dan (**) ke persamaan (***)

Karena $\text{Var}(\varepsilon) = E(\varepsilon\varepsilon^t) = \sigma^2 I$ maka matriks variansi dari $\hat{\beta}$ diperoleh

$$\text{Var}\left(\hat{\beta}\right) = (X^tX)^{-1}\sigma^2 \quad (14)$$

merupakan varians terkecil dari semua penaksir linear tak bias .

4. KESIMPULAN

Taksiran parameter yang diperoleh adalah sebagai berikut :

$$\hat{\beta} = \left(X^tZWZ^tX\right)^{-1}X^tZWZ^tY$$

5. DAFTAR PUSTAKA

[1] Agresti, Alan. 2007. "*An Introduction To Categorical Data Analysis",* Second Edition. John Wiley and Sons,Inc.,Hoboken, New Jersey.

[2] El-Habil, Abdallah M. 2012. *An Application on Multinomial Logistic Regression Model. Pakistan Journal of Statistics and Operation Research*. Vol_8. No. 2.

[3] Hall, Alastair R. 2003. *A Companion to Theoretical Econometrics*. Blackwell Publishing Ltd: *Texas A & M University*

[4] Hardiani, Erna Tri dan Amran. 2007. *Seleksi Model Multinomial Logit Melalui Akaike's Information Criterion (AIC).* Jurnal

Matematika, Statistik dan Komputasi. Vol. 4, No.1, 43-53.

[5] Matyas, Laszlo. 1999. "*Generalized Method of Moment"*. Cambridge University Press. New York.